

SYSTEMATICKÉ CHYBY PŘI ANALÝZE NELINEÁRNÍCH SYSTÉMŮ

A. Novák, F. Rund

Katedra radioelektroniky K13137, FEL ČVUT Praha

Abstrakt

Tento článek se zabývá problematikou nelineárních systémů, s nimiž se setkáváme denodenně, a z nichž mnohé považujeme pro zjednodušení za lineární. Cílem článku je upozornit na chyby vzniklé při nelineárním zpracování signálu, které nejsou na první pohled zřetelné a jež jsou velmi často zanedbávány, či v horším případě úplně opomíjeny. V článku jsou ukázány některé důsledky zanedbání nelinearity ať už z pohledu zpracování signálu a tvorby efektů v digitální doméně – v prostředí Matlab, nebo při měření zařízení jakými mohou být zesilovač, reproduktor apod. Článek je zaměřen na tři části: chyby při měření soustav s nelinearitami, jež považujeme za lineární; chyby při měření nelineárního zkruslení a chyby při samotném zpracování či tvorbě efektů. Ve všech případech mohou být buď naměřeny chybné výsledky, nebo může dojít k nesprávnému zpracování signálu. Všechny tyto negativní důsledky pak mohou z hlediska vnímání signálu působit rušivě.

1 Úvod

V přírodě, ve fyzice a samozřejmě i v akustice existuje mnoho nelineárních systémů, na něž se pro zjednodušení pokoušíme dívat jako na lineární, nicméně ve skutečnosti lineárními nejsou. Např. zesilovač, či reproduktor je obecně považován za lineární, nicméně ve své podstatě je každý zesilovač i reproduktor nelineárním prvkem. Většinou se v oblasti nízkých amplitud chovají téměř lineárně, nicméně s růstem amplitudy budících signálů jejich nelinearita roste. Takovéto systémy pak bývají občas nazývány jako systém s nelinearitami. V následujících kapitolách se pokusíme objasnit některé základní vlastnosti takovýchto systémů a pokusíme doporučit jak předejít základním chybám, jichž se můžeme ať už během měření, zpracování, či modelování dopustit.

2 Nelineární systém

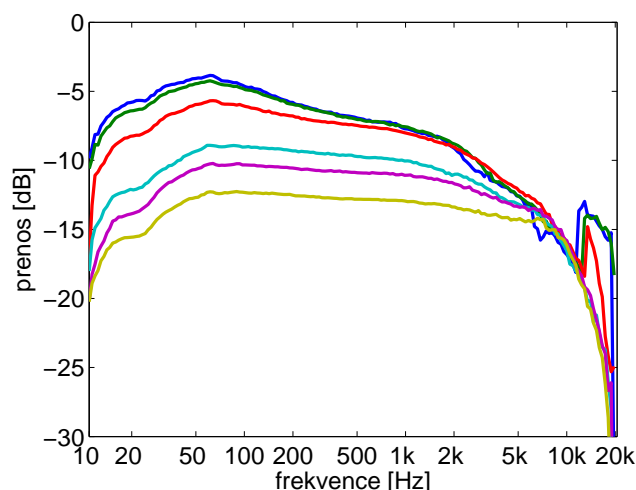
Z matematického hlediska lze nelineární systém definovat jako systém, pro který neplatí princip superpozice. Jinými slovy, pokud vstupní signál $x_1(t)$ způsobí odezvou $y_1(t)$ a jiný vstupní signál $x_2(t)$ způsobí jinou odezvou $y_2(t)$, potom odezva na součet vstupních signálů $x_1(t)$ a $x_2(t)$ nemusí být rovna součtu odezev $y_1(t)$ a $y_2(t)$. Dalším důsledkem je pak to, že je-li odezva na signál $x_1(t)$ rovna $y_1(t)$, nemusí být odezva na signál $nx_1(t)$ rovna $ny_1(t)$ ale jinému signálu $y_n(t)$, který nemusí mít ani co do tvaru, ani co do spektrálního složení žádnou podobnost se signálem $y_1(t)$. S touto vlastností (nelinearitou) se často setkáváme ve zvukové technice, kdy tzv. přebudíme audio systém.

3 Přenosová funkce a její definice v nelineárních systémech

Přenosová funkce jako taková je striktně definována pouze pro lineární systémy a v nelineárních systémech ztrácí pro svou nejednoznačnost význam. Proto přenosová funkce celého nelineárního systému není definována. Zde je třeba zdůraznit slovíčko "celého", neboť mnoho teorií o nelineárních systémech – ať už Volterrova teorie [3], teorie MISO [4, 5], nebo metoda měření s přeladovanými siny [6] používá pojmu přenosová funkce pouze pro část celého nelineárního systému a v takovém případě je výraz správně použit, neboť je v kontextu s lineární částí.

Nicméně v praxi se setkáváme s tím, že je například k zesilovači udáván v katalogovém listu činitel harmonického zkreslení, který se pokouší jedním číslem charakterizovat nelineární zkreslení. Z tohoto pohledu je tedy na zesilovač pohlíženo jako na zařízení způsobující nelineární zkreslení, tedy zařízení s nelinearitami. Zároveň je však ve stejném katalogovém listu uvedena i přenosová funkce celého zesilovače, což je z hlediska teorie o nelineárních systémech velmi zavádějící.

Je třeba si uvědomit, že se změnou vstupní amplitudy signálu se mění nejen činitel nelineárního zkreslení, ale i samotná přenosová funkce. V oblastech nízkých amplitud, kde se očekává lineárnost má přenosová funkce svůj význam, nicméně v oblasti, kde už nelineární zkreslení značně narůstá, se při změně úrovně vstupního signálu jednak změní frekvenční charakteristika a jednak též závisí na tom, jakým způsobem byla tato frekvenční charakteristika získána. Výsledkem pak může být buď přenos první harmonické - v tom lepším případě, nebo frekvenční charakteristika, jejíž hodnoty jsou zatíženy značnou chybou, kde tato chyba roste s harmonickým zkreslením. Takovýto příklad výsledků získaných měření systému v nelineárním režimu je ukázán na obrázku 1. Zde je patrné, že pro různé úrovně buzení systému (šlo o audio-zesilovač) získáváme různé přenosové charakteristiky. Následující výčet metod pro měření přenosové charakteristiky ukazuje, jakých výsledků se dopátráme, budeme-li měřit systém vykazující nelineární zkreslení.



Obrázek 1: Přenosové charakteristiky zesilovače měřené při různých úrovních zesílení v nelineárním režimu.

Metoda založená na harmonických signálech

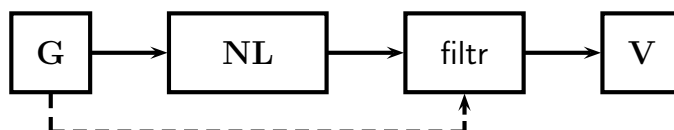
Tato metoda je založena na ručním přeladování jedné harmonické v žádaném kmitočtovém rozsahu. Je třeba pouze generátor harmonického signálu, jenž je použit jako budící signál a střídatý voltmetr, jenž na výstupu soustavy měří efektivní hodnotu. Ten kdo měření provádí, pak přeladuje ručně generátor po určitých krocích a hodnoty na výstupu, tedy z voltmetru, zapisuje. V takovém případě existují dvě možnosti. V tom horším případě je použito pouze vybavení (obr. 2a), jež bylo popsáno, a výsledky v nelineárním režimu jsou takřka nepoužitelné. V tom lepším případě je před voltmetrem umístěn filtr, či banka filtrů (obr. 2b), jenž je spojena s generátorem a plynule přeladuje kmitočet úzkopásmové propusti v závislosti na kmitočtu vstupního signálu, čímž filtruje pouze první harmonickou. Výsledkem je pak pochopitelně přenos první harmonické. Nutno však podotknout, že takovéto metody jsou spíše používány pouze pro výukové účely, kde je kladen důraz spíše na pochopení problematiky měření soustav, jež jsou považovány za lineární a v reálných měřeních se již nepoužívají.

Metoda založená na přeladovaných sinech

Tato metoda se začala více používat teprve v posledních letech a její adaptace přináší mnoho úspěchů v analýze jak lineárních tak nelineárních systémů [6, 7, 8]. Opět se zde můžeme setkat se dvěma modifikacemi. V obou případech je budícím signálem signál, jehož frekvence se neustále zvyšuje a to buď lineárně, nebo logaritmicky. V tom prvním případě se pak frekvenční charakteristika odečte z obálky výstupního signálu, což může být v případě nelineárního signálu opět zavádějící a způsobovat chyby. Výsledky zobrazené na obrázku 1, byly získány právě touto metodou. V tom druhém případě se používá tzv. inverzní filtrace, výstupní signál je nejprve podroben konvoluci s inverzním filtrem a výsledkem jsou odseparované impulsní odezvy odpovídající přenosům jednotlivých harmonických složek. Odezva první harmonické je pak použita na výpočet frekvenční charakteristiky. V tomto případě je tedy získán přenos první harmonické.



(a) obyčejná metoda



(b) metoda s filtrem svázaným s generátorem

Obrázek 2: Blokové schéma měření přenosových charakteristik, G - generátor, NL - nelineární systém, V - voltmetr

MLS (Maximum length sequences) metoda

Tato metoda se stala velmi oblíbenou na konci minulého století, nicméně v dnešní době je již vytlačována právě zmíněnou metodou používající přeladované siny. Vstupním signálem této metody je pseudonáhodná posloupnost, jež má takové vlastnosti, že se ze statistického hlediska chová jako bílý šum. Po cyklické korelaci s výstupním signálem získáváme přímo impulsovou odezvu a z ní pak již snadno získáme přenosovou funkci. Problém opět nastane v případě, kdy je analyzován nelineární systém. Proces korelace rozmělní nelineární složky do impulsové odezvy jako mnohačetné vrcholy, které ani z impulsové odezvy, ani z následné frekvenční charakteristiky zatím není možno odstranit. Tedy opět získáváme frekvenční charakteristiku zatíženou chybou.

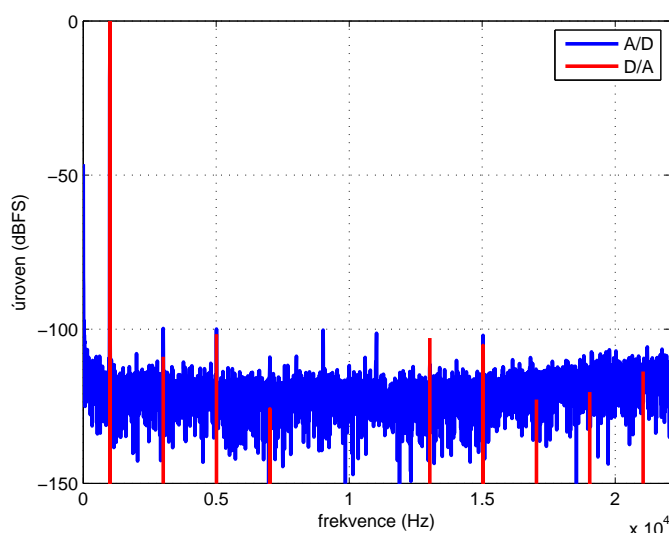
Na další problémy, které znesnadňují použití této metody jsme upozornili článkem [1].

4 Chyby při měření nelinearit

Pokud při měření nelinearit použijeme některou z metod, která vyžaduje číslicové zpracování signálu, což je případ druhých dvou výše zmíněných metod, musíme si uvědomit, že i toto měření má své specifika. Signál je číslicově generován v počítači (např. pomocí Matlabu), převeden na analogový signál pomocí zvukové karty (D/A převodníku), dále prochází měřeným systémem, zvukovou kartou (A/D převodníkem) je zpět převeden na číslicový signál který je pak číslicově vyhodnocován. V prostředí Matlab lze číslicově generovaný signál uložit ve formátu

.wav příkazem `wavwrite`, tento přehrát na zvukové kartě, a po průchodu měřenou soustavou ho externím software opět uložit ve formátu .wav a načíst jej do Matlabu příkazem `wavread`. Dále se nabízí možnost využít pro generování i příjem signálu Data Acquisition Toolbox, který ovšem nepodporuje vysoké vzorkovací kmitočty a kvantovací hloubky, podobně jako příkazy `wavplay` a `wavrecord`. Je zřejmé, že hlavní problém spočívá v převodu z digitálního signálu na analogový a zpět. Při měření metodou MLS navíc dochází i k problémům při nesynchronosti vzorkovacího kmitočtu A/D a D/A převodníku, jak bylo ukázáno v [1].

Při měření s harmonickým signálem je nutné si naopak uvědomit, že spektrum kvantovaného harmonického signálu neobsahuje pouze základní harmonickou složku, ale i intermodulační produkty, jako jsou součtové a rozdílové násobky vstupních kmitočtů, viz např. [2]. Velikost těchto produktů je úměrná velikosti kvantovacího kroku, tedy nepřímo úměrná 2^N , kde N je použitý počet bitů. Je nutné upozornit, že jde o skutečně využitý počet bitů, tedy při plném rozkmitu signálu. Ke vzniku těchto produktů dochází při generování číslicového signálu s odpovídající bitovou hloubkou i při A/D převodu výstupního signálu ze soustavy. Pro ilustraci viz obrázek 3, který uvádí příklad spektra signálu o jedné kmitočtové složce po kvantizaci (červeně) a po průchodu analogovým obvodem a A/D převodu (modře). Z uvedeného plyne, že měřené zkreslení pochází nejen z měřené soustavy, ale podílí se na něm i D/A a A/D převod. Pro minimalizaci tohoto zkreslení je nutné používat zvukové karty s co největším počtem bitů (a celé zpracování musí probíhat s velkou bitovou hloubkou) a oba převodníky zvukové karty musí být maximálně vybuzené.



Obrázek 3: Spektrum číslicového signálu (16 bit) o jedné harmonické složce, červeně číslicově generovaný signál, modře tento signál po průchodu analogovým obvodem a A/D převodu.

5 Aliasing při nelineárním zpracování

Zpracováváme-li signály v digitální podobě a pracujeme-li s nimi, musíme v první řadě zamezit tzv. aliasingu, jenž úzce souvisí s pojmy digitalizace a vzorkovací frekvence. Chceme-li nějaký analogový signál převést na digitální, je třeba jej lehce upravit, aby nedošlo k právě zmíněnému efektu. Frekvenční složky vyšší než polovina vzorkovacího kmitočtu musí být před samotným procesem vzorkování ze signálu odstraněny, neboli odfiltrovány. Nebyly-li by odstraněny, mohlo by dojít k nepříjemnému zrcadlení složek vyšších než je polovina vzorkovacího kmitočtu do pásma pod polovinou vzorkovací frekvence. Tento jev je v číslicovém zpracování považován za základní znalost a při následujícím lineárním zpracování už na něj není třeba pomýšlet. Nicméně máme-li už signál správně digitalizovaný a aplikujeme-li na něj nelineární operace, může opět dojít k jevu zvanému aliasing. Artefakty způsobené tímto jevem se přitom už následnou filtrací

odstranit nedají. Jediné známe řešení problému je tak pracovat s dostatečně vysokým vzorkovacím kmitočtem, a s nelinearitami jejichž řád je dostatečně malý. Provádíme-li pak operaci typu tvrdý limiter – tzv. saturace, dojde k aliazingu v každém případě, neboť má tato nelineární funkce nekonečný rozvoj, tedy její řád není limitován.

Proč k aliazingu při nelineárním zpracování dochází, se pokusíme objasnit na následujícím příkladu doplněném sérií grafů získaných v prostředí Matlab. Ty porovnávají aplikaci nelineární funkce na harmonický signál v analogové doméně, kde se s jevem aliazing nesetkáme a v digitální doméně, kde naopak ano.

Vstupním signálem pro obě ukázky je harmonický signál o frekvenci 2500 Hz. Vzorkovací kmitočet pro digitalizaci je 8000 Hz, tedy vstupní signál leží v tomto případě pod polovinou vzorkovacího kmitočtu, nicméně nad jeho čtvrtinou. Pro lepší názornost bude frekvenční osa vynášena v radiánech, tak jak se občas vynáší v digitálním zpracování, aby byl patrnější vztah mezi vzorkovacím kmitočtem. V takovém případě pak π odpovídá jeho polovině a tudíž 2π je právě rovno vzorkovacímu kmitočtu. Na obrázku 4 je tedy vynesena posloupnost celého zpracování, vlevo digitální a vpravo analogové.

Na obrázcích 4a a 4b jsou vyneseny spektra vstupních signálů (v tomto případě jsou oba totožné, neboť jak pro analogovou, tak pro digitální ukázku používáme tentýž signál). Na obrázcích 4c a 4d jsou pak tyto grafy rozšířeny na oblast až do dvojnásobku vzorkovacího kmitočtu, aby byl patrný rozdíl mezi spektrem digitálního (4c) a analogového (4d) signálu. Na obrázcích 4e a 4f je pak zobrazeno spektrum signálů, na něž byla aplikována nelineární funkce. V obou případech šlo o operaci mocniny dvou, tedy je-li vstupní signál $x(t)$ a výstupní $y(t)$, můžeme jejich vzájemný vztah zapsat jako $y(t) = x^2(t)$. Aplikujeme-li takovouto nelineární operaci na harmonický signál, měla by být výsledkem stejnosměrná složka a druhý harmonická, obě s poloviční amplitudou. Je třeba si však uvědomit, že spektrum signálu je chápáno jako komplexní a tudíž je druhá harmonická rozdělena na svou kladnou a zápornou frekvenční část a proto má při vynesení pouze kladné frekvenční osy velikost o polovinu nižší. Ze stejného důvodu je i vstupní harmonický signál zobrazen pouze s poloviční amplitudou. Na obrázku 4f je patrné, že druhá harmonická se nám dostala nad π tedy nad polovinu vzorkovacího kmitočtu. V analogové doméně se pochopitelně neděje nic, ovšem v digitální (obr. 4e) dojde k zrcadlení tohoto kmitočtu zpět kolem π a výsledkem je harmonická složka mezi $\pi/2$ a π , která vznikla pouze jako produkt aliazingu a je tedy nežádoucí. Poslední dva obrázky 4g a 4h pak znázorňují oblast do π pomocí zelené čerchované charakteristiky znázorňující filtr, jenž vybírá skutečný signál. Je zde patrné, že se tyto dvě spektra obou signálů liší, a s nežádoucí aliazingovou složkou v digitální doméně (obr 4g) už nic neuděláme. V analogové doméně by nám při případné digitalizaci naznačeným čerchovaným filtrem došlo sice k odfiltrování složky vyšší než π , nicméně je to lepší situace, než když jsou ve výsledku obsaženy jiné složky, než které jsme očekávali a které do výsledku patří. V případě komplexnějšího signálu ani nemusíme podle spektra poznat, zda k aliazingu vůbec došlo, natož které složky jsou žádoucí a které ne.

Jak již bylo zmíněno výše, s aliazingem při digitálním zpracování těžko něco dělat. Problém lze pouze snížit volbou vysokého vzorkovacího kmitočtu. Úplného odstranění bychom dosáhli pouze použitím takových nelineárních funkcí, jejichž řád by byl s ohledem na maximální kmitočet volen tak, aby nedocházelo k aliazingu. Již zmíněný tvrdý limiter by musel být aproximován funkcí s takovým rozvojem, aby maximální kmitočet po aplikaci nelineární funkce nebyl vyšší než polovina kmitočtu vzorkovacího. V mnoha aplikacích by takové řešení vedlo pouze ke komplikacím, ale je dobré si uvědomit, že veškeré nelineární efekty aplikované v dnešním digitálním světě mohou vést k jiným výsledkům, než očekáváme. Důvodem pak může být právě popsany aliazing způsobený nelineárním zpracováním.

6 Závěr

Tento článek shrnuje možné systematické chyby, se kterými se lze setkat při práci s nelineárními systémy. Je nutné vzít do úvahy, že mnoho systémů, ke kterým se přistupuje jako k lineárním, jsou ve své podstatě nelineární. V případě nelineárního systému není definována přenosová funkce systému jako celku, proto je vždy třeba uvažovat, čemu naměřená charakteristika odpovídá a vyvarovat se výše popsaných chyb zejména vhodnou volbou měřicí metody.

Pokud, jak je v dnešní době časté, pro generování měřicího signálu i zpracování signálu naměřeného využíváme číslicového zpracování, musíme si uvědomit, že přecházíme z diskrétní do analogové domény a zpět, a že tím do měřené soustavy zařazujeme další nelineární prvky. Toto je významné zejména při měření pomocí harmonického signálu, kdy dochází ke vzniku intermodulačních produktů. Proto je nutné pracovat s dostatečným bitovým rozlišením.

Poslední problematikou, kterou je nutné nepřehlížet při práci s nelineárním systémem, je problematika aliažingu. Pokud prochází signál nelineárním systémem (a tím je i samotné kvantování), dojde ke vzniku vyšších kmitočtových složek, a tudíž může dojít k aliažingu. Taktéž samotné nelineární zpracování v digitální doméně (např. kompresor dynamiky, limitér apod.) může do signálu vnést nežádoucí složky důsledkem aliažingu. Pokud dojde k výše popsanému aliažingu uvnitř zpracování diskrétního signálu, není možné na obecné úrovni tento jev odstranit. Pro zmírnění těchto jevů je nutné pracovat s dostatečně vysokým vzorkovacím kmitočtem, nebo pro úplné potlačení používat nelineární modely s omezeným stupněm nelinearity.

Závěrem lze tedy říci, že číslicové zpracování signálu přináší sice mnoho výhod, ale při jeho využití pro měřicí účely je nutné pro minimalizaci výše specifikovaných chyb pracovat s velkou bitovou hloubkou a vysokým vzorkovacím kmitočtem. Pracujeme-li pak s nelineárními systémy v digitální doméně (digitální efekty apod.) je třeba pracovat obezřetně a vyvarovat se chyb pocházejících z podstaty nelineárního zpracování signálů.

7 Poděkování

Popisovaný výzkum je podporován grantem GAČR 102/05/2054 – Kvalitativní aspekty zpracování audiovizuální informace v multimediálních systémech.

Reference

- [1] Novák, A., Rund, F.: Problém špatné synchronizace vzorkovacích kmitočtů u MLS signálů: Model v prostředí MATLAB, in *Technical Computing Prague 2006*, Praha, 2006.
- [2] Kadlec, F.: *Zpracování akustických signálů*, Praha, FEL ČVUT, 2002.
- [3] Schetzen, M.: *The Volterra and Wiener theories of nonlinear systems*, New York, John Wiley & Sons, 1980.
- [4] Bendat, J. S.: *Nonlinear System Techniques and Applications*, New York, John Wiley & Sons, 1998.
- [5] Marchesiello, S.: Application of the Conditioned Reverse Path Method, *Mechanical Systems and Signal Processing 17(1)*, 2003, p. 183-188.
- [6] Farina, A.: Simultaneous measurement of impulse response and distortion with a swept-sine technique, in *Proc. 108th AES Convention*, Paris, February, 2000.
- [7] Kite, T.: Measurement of Audio Equipment with Log-Swept Sine Chirps”, in *Proc. 117th AES Convention*, San Francisco, October, 2004.

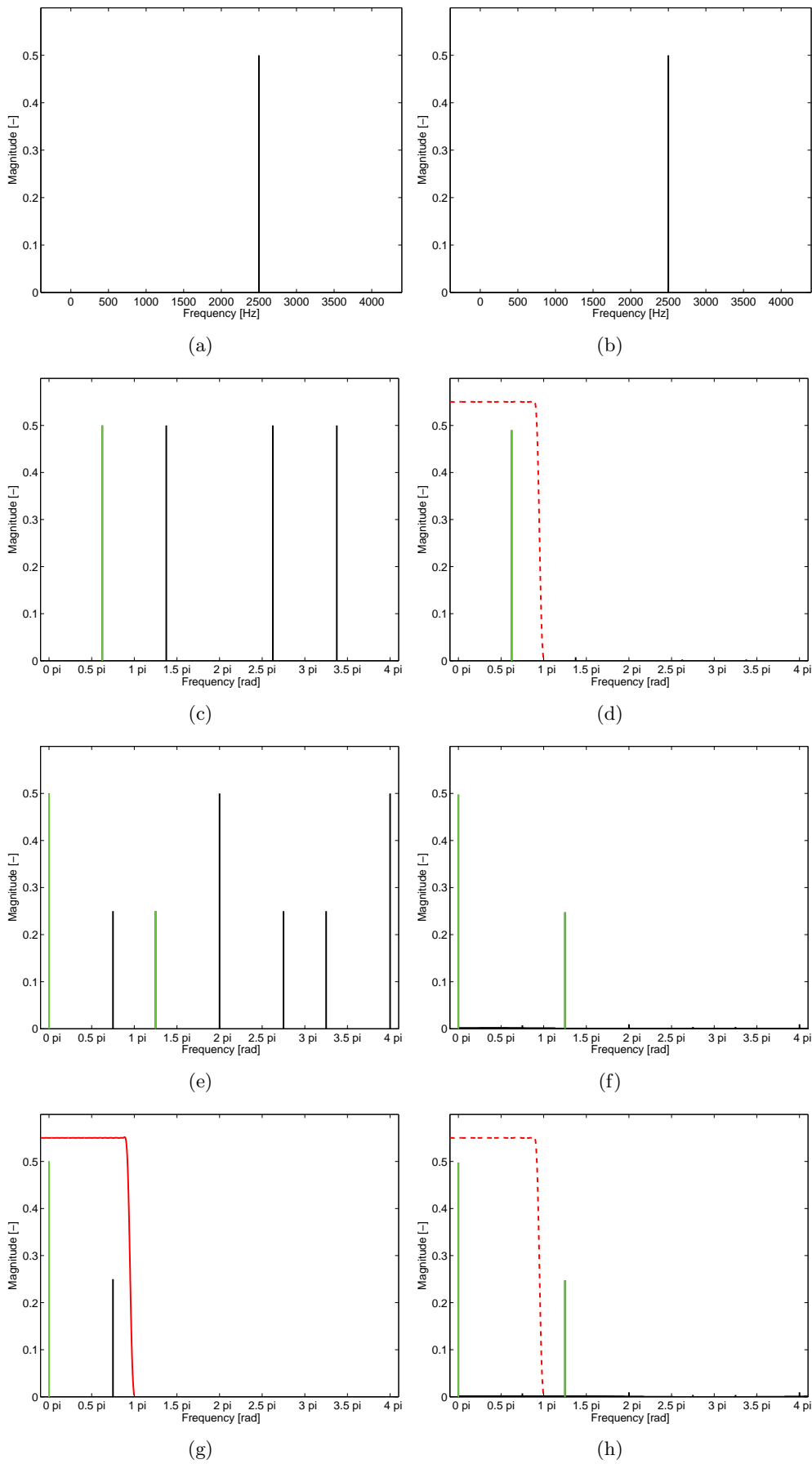
- [8] Müller, S., Massarani, P.: Transfer function measurement with sweeps *J. Audio Eng. Soc.*, Vol 49, N° 6, June, 2001.
-

Antonín Novák

Katedra radioelektroniky, FEL ČVUT Praha, Technická 2, 166 27, Praha 6
e-mail: novaka4@feld.cvut.cz

František Rund

Katedra radioelektroniky, FEL ČVUT Praha, Technická 2, 166 27, Praha 6
tel. +420 22435 2108, e-mail: xrund@feld.cvut.cz



Obrázek 4: Spektrum signálů při nelineárním zpracování signálu, vstupní signál o frekvenci $f = 2500$ Hz, nelineární funkce $y(t) = x^2(t)$, digitální zpracování vlevo, analogové vpravo